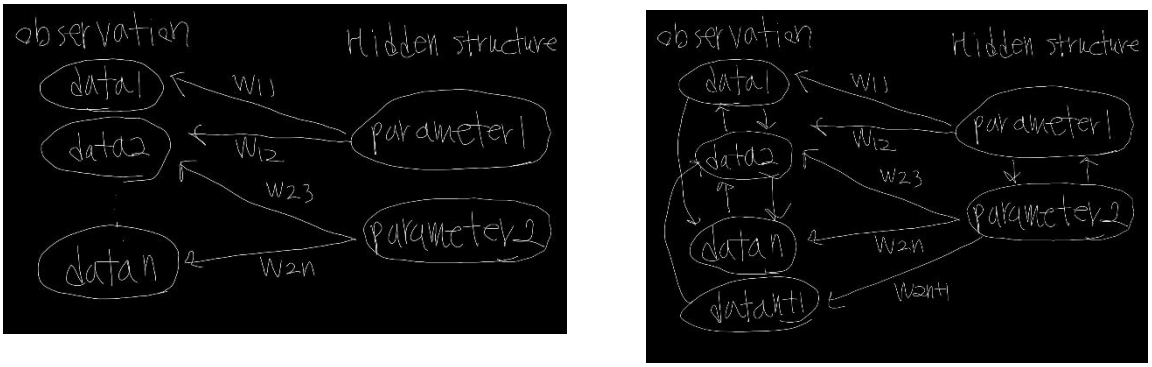
X: observed data

H: hidden variable

W: weight matrix

There are many situation which we want to use small number of models for explaining and predicting observation data. Principle component analysis(PCA), Factor Analysis(FA) are good example of it. The left picture in Figure1 shows that independence assumption between observation must be satisfied and other variables which effect the data except model does not exist. The former combined with Gauss-Markov theorem makes many algorithms use Frobenius norm for optimization. And the letter come from the fact that we assume that hidden model explain completely data like regression.

In addition, some algorithms like PCA or FA assume the independence between hidden variables, others like NMF does not. When we use the result of dimension results, we should consider direction of research because the correlations between results depend on algorithms. For example, in neural science, people does not use PCA for dimension reduction because they use them for making network model, which we can't use PCA in this situation because PCA produce the orthogonal results.



<figure1: the graphical structure assumption for estimation algorithms. Many ones use the left assumptions, so we need to modify algorithms if the right structure must be used>

Also, we need to check distribution of hidden variables because the Frobenius norm produce the result which are same when the variable follows normal. We can determine the distribution by drawing plot of data or comparing momentums. If this is not normal, then we need to use Maximum Likelihood estimation(MLE) for efficiency. However, the computation cost is too big to implement, so we need to use other method like Monte-Carlo or MCMC else.

Then, how can we fit hidden model for data which have correlation each other. One way is to use two least square estimation(2LSE) which is a method using weight matrix made by error of Ordinary Least Square(OLS) regression. That is Generalized Least Square(GLS) with using covariance matrix of

residue of OLS error as weight. We can use this because the OLS does not catch the correlation between observation, so we can think the information about correlation is in the residue of OLS. I think we can use this method in MLE estimation by reflecting covariance matrix between observations in joint distribution.